

Identifying UMLS Concepts in Emergency Department Terms



using
Domain Knowledge and
Natural Language Processing Techniques

Debbie Travers

July 27, 2001

Mentor

- Olivier Bodenreider

Summer Program

- May Cheh
- Tom Rindflesch

Clinical Data

University of North Carolina Hospitals
Cape Fear NC 27514

EMERGENCY DEPARTMENT NURSING RECORD

Date 04/07/01 Time 15:06

MR. #80

[REDACTED] 33 Y M

AMB. SKIN W/B. LYO LACERATION
IN A ABOVE RIGHT EYE. BUTTED HE AD WITH SOMEONE
ELSE WHILE P LAYING SOCCER- DENIES LOC, B
LEADING CONTROLLED. ICE ON W
D. BLURRED VISION

CHIEF COMPLAINT LACERATION

36.7 138 078 083 18

VS TEMP 37 P RR 18

TRACHEA CORD TENDON PEDANTHESES

PHASE 1

ALLERGIES NONE

REFERRAL AGENT AGENCY

Plan

Contin

Standards

- DEEDS

<http://www.cdc.gov/ncipc/pub-res/pdf/deeds.pdf>

- Data element 4.06 "Chief Complaint"

- Emergency Department
Chief Complaint Thesaurus project in NC

Chief Complaints- No Standards

- fever, fevers, temp 103.6
- vomitting, vomiting, emesis, dry heaves
- bug in ear, FB ear
- foot inj, foot injury, LE injury, hurt foot
- tvfellowhim

Today's Presentation

- Project description, methods
- Research results
- Future directions

Project Description

Objective

To apply and evaluate methods for processing Emergency Department (ED) Chief Complaint (CC) terms, in order to identify the concepts that comprise the ED CC domain.

Schoeffler, Travers, Hales, Waller & Tintinalli

Evaluating Emergency Department Chief Complaint Data
NLM Trainees Meeting, 1999

- Results
 - 14% match rate with UMLS concepts
- Conclusion: additional processing needed
 - Multiple concepts per entry
 - Punctuation

Current Study- Goals

- Improve match rate with UMLS concepts
- Use domain knowledge to identify patterns
- Apply & evaluate NLP methods for handling ED text
- Pilot project which will later apply to larger sample

Materials

- Expand to 3 hospitals
 - Rural, urban, suburban
- All ED visits for one month
- Data characteristics
 - 12-character limitation, UNC
 - Controlled list, CMC

Materials Preparation

ED CC Terms
N= 6365



Exclude
Pregnancy Cases
n=6289



Eliminate case,
Leading spaces, "?", etc.
n=6054

Results

M

```
graph TD; M((M)) --> Match[Match<br/>n=1327<br/>(22%)] ; M --> NoMatch[No Match<br/>N=4727<br/>(78%)]
```

Match
n=1327
(22%)

No Match
N=4727
(78%)

Methods

- Punctuation Processing
- Expand
 - Acronyms, abbreviations, truncations
- Delete
 - Modifiers, qualifiers

Punctuation Processing

Chief Complaints with Slash, Comma

- **Replace**

- Common abbreviations, acronyms:

h/a

headache

b/p elevated

blood pressure elevated

- **Expand**

- Coordinate constructions:

hip/thigh pain

hip pain

thigh pain

- **Retain Body Location**

- Signs, symptoms, injuries:

laceration/forehead

forehead laceration

Punctuation Processing

Chief Complaints with Slash, Comma

- **Eliminate**

- Unnecessary abbreviations:

c/o earache

earache

- **Delete slash, comma, semi-colon**

- & split into 2 terms:

fall/rib pain

fall

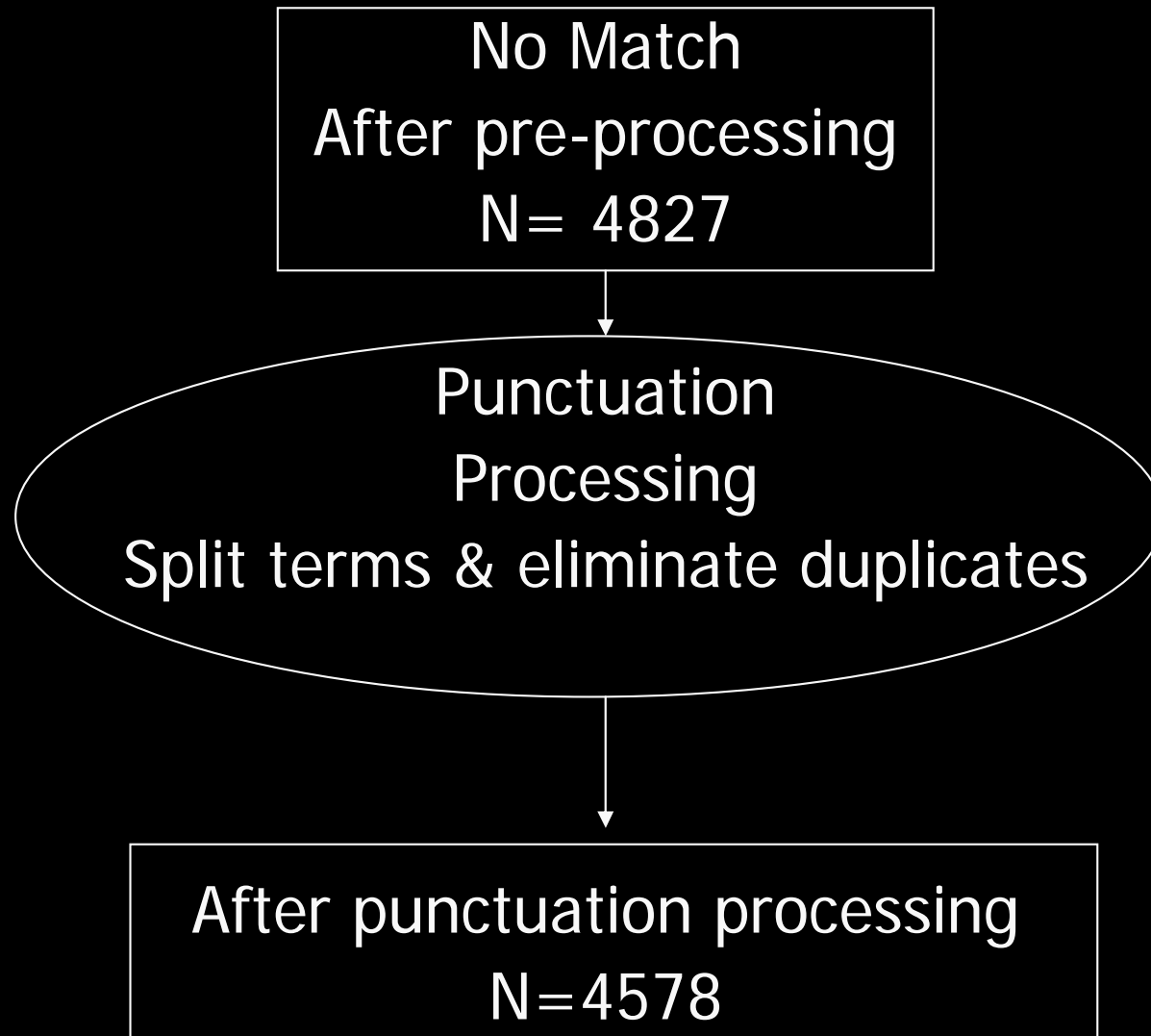
rib pain

Earache, sore throat

earache

sore throat

Punctuation Processing



After Punctuation-processing

M



```
graph TD; M((M)) --> Match[Match<br/>n=595<br/>(13%)]
M --> NoMatch[No Match<br/>N=3983<br/>(78%)]
```

Match
n=595
(13%)

No Match
N=3983
(78%)

Expansion

Acronyms, abbreviations, truncations

- Segmented into words
- Examined word frequencies
- Expanded frequently-used acronyms, abbreviations, truncations:

FB	foreign body
Rx	reaction (if after all, allerg, else prescription)
Diarr	diarrhea

After Expansion

M

```
graph TD; M((M)) --> Match[Match<br/>n=358<br/>(9%)]<br/>NoMatch[No Match<br/>N=3625<br/>(91%)];
```

Match
n=358
(9%)

No Match
N=3625
(91%)

Deletion

Modifiers, qualifiers

- List of modifiers, qualifiers from literature
- Examined non-matches, calculated frequencies
- Deleted frequently-used modifiers, qualifiers:

Modifier

Right leg injury

Severe chest pain

Qualifier

History of seizure

Headache since 5 am

After Deletion

M

```
graph TD; M((M)) --> Match[Match<br/>n=666<br/>(18%)]
M --> NoMatch[No Match<br/>N=2959<br/>(82%)]
```

Match
n=666
(18%)

No Match
N=2959
(82%)

Results Summary

- 22% match rate, UMLS tools
 - 35% match rate, custom tools
-
- 2946 CC terms (1855 unique) matched UMLS concept(s)
 - Single concept- 68%
 - Multiple concepts- 32%

Evaluation- Matches

- Single concept match
 - manual review
 - 8% inaccurate
 - Stepped on by sibling|Step sibling|C0337504
- Multiple concepts
 - semantic groups
 - 15% not from same semantic group
 - Bipolar (Disorder, Disease or syndrome) vs.
 - Bipolar (Concepts & Ideas, Spatial concept)

Non-matches

- **Misspellings**

abcess, assualt, sycopy

- **Acronyms, abbreviations, truncations**

tachy, ASD, 7273, ED D# pt

- **Meta-information**

diagnosis meningitis, confusion

Non-matches

- **Un-mappable**

1 week, 12/14/00, a bleed, car came through apartment

- **Lumps, bumps, knots on/to**

lump on head, knot on thigh, bump to face

- **?not in Metathesaurus**

motor vehicle crash- rollover, re-check cellulitis

Conclusions

- Domain knowledge supplemented with
 - Simple pattern-matching approach
- More risk
- CC terms finer-grained than terminologies
 - Granular terms broadened
- Manual evaluation of accuracy of matches will be necessary when building the ED CC thesaurus

Future Directions

- Continue to refine methods
 - Patterns, non-matches
- Matches with the UMLS concepts
 - Pursue automated disambiguation
- Apply to larger sample
 - 12 months

Future Directions

- Provide UMLS researchers with:
 - List of words not in the Lexicon
 - List of pairs of ED entry terms/UMLS concepts
- Analyze matched terms
 - Source vocabularies
 - Semantic groups
 - Begin to build CC terminology
- Doctoral dissertation research